Novel Nonlinear and RR-based Methods for Inappropriate ICD Therapy Reduction

A THESIS

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Samuel Newell

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Alena Talkachova, Advisor

May 2021

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my advisor, Alena Talkachova, for challenging me over the last two and a half years. Learning how to do proper research is not easy, and she asked many difficult questions to develop me as a researcher. However, I always knew she had my best interests at heart, and I always knew I could ask her for support whenever I needed it.

To everyone in the Talkachova lab, thank you for your support over the last few years. Your feedback on my presentations, manuscripts and other research has been vital in my development as a researcher. It was my pleasure to have worked with each one of you.

Lastly, to my parents and family, thank you for supporting my decision to pursue my education further. I am fortunate that you have chosen to support me along my academic journey and for this I am grateful. I ask for your continued support as my educational journey continues with Medical School in August 2020.

ABSTRACT

Implantable cardioverter-defibrillators (ICD) are a commonly implanted device used to deliver arrythmia terminating therapy to the heart in the presence of life-threatening arrythmias such as ventricular fibrillation and ventricular tachycardia. However, not all therapies, shocks or anti-tachycardia pacing (ATP), are appropriate. In many instances, therapy is delivered when the heart is in an abnormal, but non-life-threatening arrythmia such as atrial fibrillation. While major medical device companies have devised numerous strategies to eliminate these cases of inappropriate therapy, there remains room for improvement. Thus, the goal of this thesis is to describe the development of novel strategies to discriminate between appropriate and inappropriate ICD therapy. The strategies use nonlinear measures and RR interval measures fed into principal component analysis or linear combination scores to discriminate. The final method using linear combination scores showed near 100% discrimination between appropriate and inappropriate therapy events retrospectively and 100% discrimination in a pseudo real time study. Thus, this strategy shows immense promise for use in a future large clinical study to eliminate inappropriate therapy.

Table of Contents

List of Tables

List of Figures

Chapter 1 – Introduction

1.1 – ICD Statistics and Inappropriate Therapy

Roughly 150,000 implantable cardioverter-defibrillators (ICD) are implanted in the United States annually [1]. These life saving devices deliver therapy to the heart in the form of anti-tachycardia pacing (ATP) or high energy shocks during life threatening cardiac arrhythmias such as ventricular fibrillation (VF) and ventricular tachycardia (VT). Thus, physicians world-wide often opt to implant them in patients deemed high risk for developing these life-threatening arrythmias. In fact, multiple studies have shown that ICDs decrease all-cause mortality in heart failure patients [2, 3].

Despite the clear beneficial nature of ICDs, there are many drawbacks associated with them, mainly infection, cost and inappropriate therapy. Inappropriate therapy is any delivery of therapy when the heart is not in a life-threatening arrythmia such as VT/VF. With transvenous (TV)-ICDs it most commonly occurs due to non-life threatening arrythmias originating from the atria, known as supraventricular arrythmias. Among the most common of these, atrial fibrillation (AF) is the leading reason of inappropriate therapy for TV-ICD devices [4]. Given AF is an epidemic itself, affecting millions of Americans [5], this is not surprising. Other common reasons for inappropriate therapy for TV-ICDs includes and is not limited to supraventricular tachycardia (SVT) and oversensing due to over-sensed cardiac signals, external signals or device malfunction such as lead fracture. In the newest subcutaneous (S)-ICD devices oversensing is the primary reason for inappropriate therapy [6].

Inappropriate shocks had a 1 year incidence of 2.4% and 2.5% cases of therapy in the newest subcutaneous (S)-ICD and single lead TV-ICD devices respectively [7, 8] (see Figure 1). Inappropriate therapy (shocks and ATP) had a 1 year incidence of 3.4% in the single lead TV-ICD devices [7]. While these numbers are significant improvements over past statistics due to improved detection algorithms and therapy delivery strategies, they disguise the problem as 14% of shocks were inappropriate in the recent PainFree study [7]. This is of major concern because multiple studies have shown that high rates of inappropriate shocks lead to lower quality of life and increased morbidity and mortality

[9, 10]. So, while progress has been made, there still is room for improvement in lowering the number of inappropriate therapies.

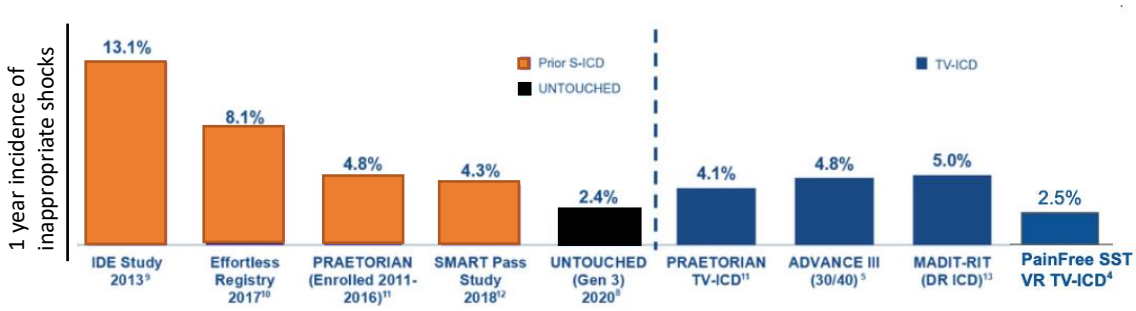


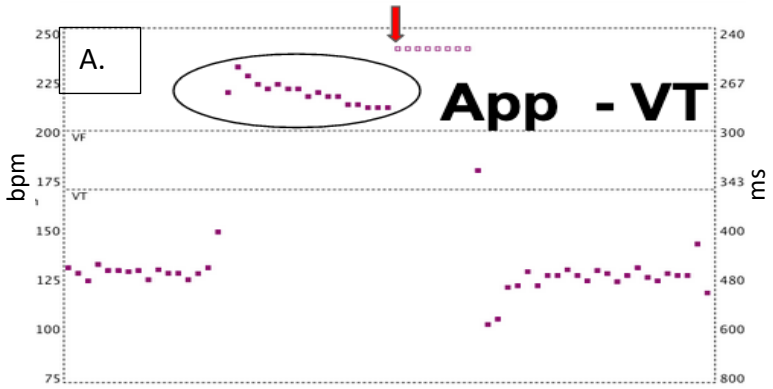
Figure 1: Recent clinical studies showing 1 year incidence of inappropriate shocks. Columns 1-8 adapted from Boston Scientific [8] and augmented with the last column. The last columns data is taken from the PainFree trial of a Medtronic device to demonstrate newest Medtronic data [7].

1.2 – Current Methods for Therapy

Deciding when to deliver therapy is a difficult task as evidenced by the number of inappropriate therapies delivered by current devices. One wants to ensure all life-threatening arrythmias have therapy delivered to them while minimizing the amount of inappropriate therapy. However, current ICDs just rely primarily upon the rate of the tachycardia for making the decision to deliver therapy. Medtronic and Boston Scientific (BSci) TV-ICDs do so by detecting if a certain percentage of RR intervals in a window are less than a threshold time length set by the physician. Recent clinical studies have set this threshold near 350 ms for VT and 320 ms for VF in their primary prevention patients [4, 7]. Traditional windows range from 10 to 24 RR intervals in length, but the PainFree Study demonstrated that windows of 40 RR intervals were effective at reducing the number of inappropriate therapies [7]. After an initial decision to initiate therapy is made, ATP is delivered, unless it has been programmed off or the tachycardia rate is below a programmed threshold time length. However, the TV-ICD will continue to monitor the percentage of RR intervals less than the threshold in the window after ATP is delivered and will begin charging. Unless an exit condition is met because a percentage of RR intervals in the window are now above the threshold, the device will shock the heart. In

Figure 2 we see two examples of TV-ICDs making an appropriate decision to deliver ATP and one example of a TV-ICD making an inappropriate decision to deliver ATP as well as a shock. S-ICDs rely upon a slightly different strategy. When the average of the last 4 RR intervals is below the set VT/VF threshold the device enters into a tachycardia state and stays in this state unless 24 consecutive beats of at least 40 ms greater than the threshold occur [11]. If the device does not exit the tachycardia state within a desired duration, then therapy is initiated.

Beyond rate-based decision making, a variety of strategies have been developed in the last few decades to reduce the amount of inappropriate therapy. They typically look at either the stability, onset and morphology of the signal or utilize a variety of filters to lessen oversensing. Stability analysis looks at the regularity of the RR intervals because SVT/AF often has more irregular RR intervals than VF/VT. Onset analysis determines if the heart rhythm suddenly changed or gradually changed as SVT/AF often onset more gradually than VT/VF. Lastly, morphology analysis uses wavelets and vector analysis to determine if the electrocardiogram (EGM) signal looks like the patients baseline rhythm signal or a new rhythms signal by comparing the EGM against baseline QRS templates. However, all of these enhancements can be programmed off by the physician and only function in what's commonly referred to as the conditional rate zone: the region below the VF threshold and above the VT threshold, typically 300-350 ms. Thus, they often fail to eliminate AF/SVT rhythms either due to the device programming, failure to recognize an AF/SVT in the conditional rate zone or because the rate is below the conditional rate zone.



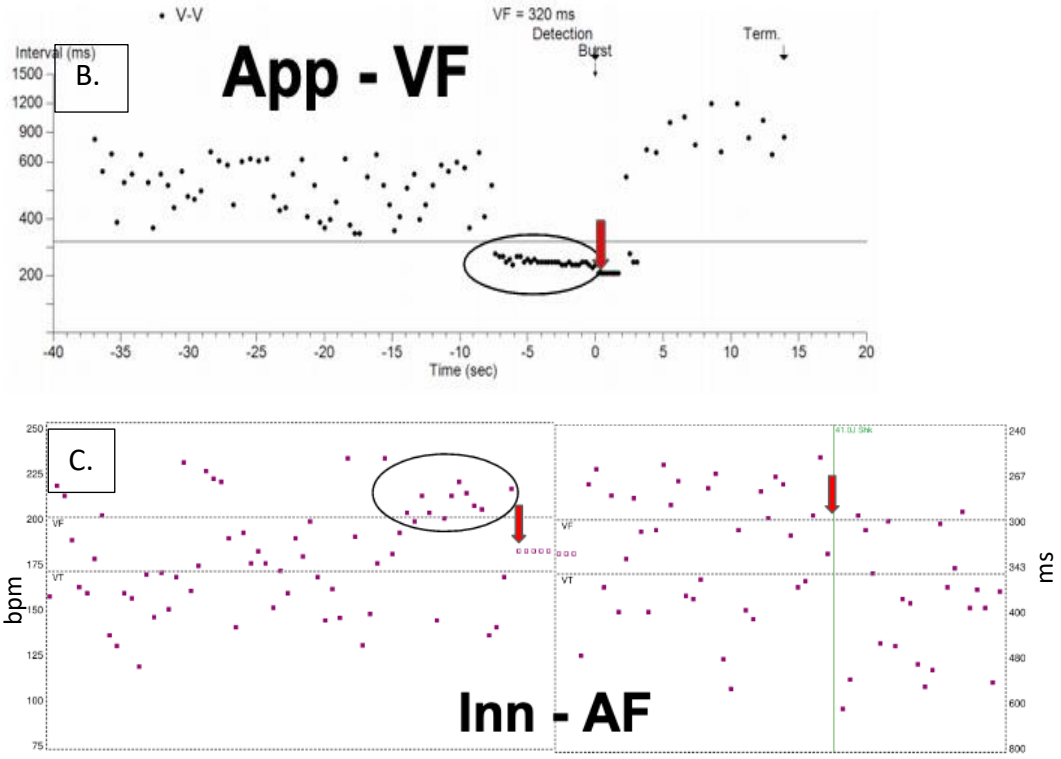


Figure 2: A (Top): A plot of RR intervals taken from the case report of a Boston Scientific (BSci) single chamber TV-ICD, the Dynagen EL ICD D150 device, demonstrating an App therapy for a rhythm classified as VT after adjudication by a cardiologist. The black oval demonstrates 17 RR intervals below the VF threshold of 300 ms prior to therapy delivery in the form of ATP at the red arrow. **B (Middle):** RR interval plot taken from the case report of a Medtronic single chamber TV-ICD, the Evera XT VR. The rhythm identified by the cardiologist was VF in this instance and thus the therapy deemed App. The black oval encompasses 30 RR intervals below the VF threshold of 300 ms prior to ATP therapy at the red arrow. **C (Bottom):** Another plot of RR intervals taken from the case report of the same device type as A. However, the rhythm was classified as AF by the cardiologist and thus the therapy in the form of ATP is deemed Inn. The black oval shows 11 RR intervals below the VF threshold of 300 ms prior to ATP therapy at the red arrow. The next red arrow indicates the decision to then shock the heart with 41.0 J.

1.3 – Motivation & Research Questions

While inappropriate therapy reduction strategies have improved in recent years, there is still a need to improve the therapy decision algorithms in order to reduce the number of inappropriate therapies, primarily inappropriate shocks. Thus, the goal of this thesis was to develop novel strategies for discrimination between inappropriate and appropriate therapy. The basis of the strategies was to use 4 novel nonlinear metrics: multiscale frequency, multiscale entropy, shannon entropy and kurtosis, along with 7 standard RR based metrics: MeanRR, pnn50, RMS, STD, SD1, SD2 and SD1/SD2, fed into either principal component analysis or linear combination scores to discriminate between inappropriate and appropriate therapy. To develop the algorithm beyond this basis, the thesis addressed these 5 primary research questions:

1) Because this study was retrospective, the signal window length could be arbitrarily chosen given enough signal was available. However, a short window will have a more stationary signal, while a long window may provide more information. Thus, an appropriate window length for discrimination needed to be determined. As such the first research question asked was: Does a long or short RR window prior to therapy lead to better retrospective discrimination?

2) An appropriate therapy could be in response to either VF or VT while inappropriate therapy was almost always in response to AF in our dataset. Thus, it was desired to know if the novel strategies could discriminate therapy in response to VF and VT from therapy in response to AF or only one or the other. Manufacturer discrepancies in "out of the box" programming could also result in differences in the ability of the novel strategies to discriminate appropriate and inappropriate therapy. Thus, the second research question asked was: Does arrythmia type and/or manufacturer influence discrimination?

3) The results of research question two demonstrated that the manufacturer influenced our ability to discriminate between appropriate and inappropriate therapy. Thus, if the dataset was split into two separate datasets, it could improve the ability of the strategies to discriminate between appropriate and inappropriate therapy. The third research then became: Does separating the data by manufacturer into two separate datasets improve discrimination?

4) In answering research questions 1-3, principal component analysis was utilized. However, simpler, linear combination scores i.e. summations of the 11 metrics, may also prove sufficient for discrimination. The advantage of the linear combination score was there was a small amount of supervision used to choose which parameters to sum as only statistically significant parameters were used in the formation of them. Thus, to prove that the linear combination scores, titled "therapy scores," were equally sufficient at retrospectively discriminating appropriate and inappropriate therapy the fourth research question asked was: Do the therapy scores discriminate appropriate and inappropriate therapy?

5) Research questions 1-4 addressed if the novel strategies could discriminate retrospectively. However, the aforementioned rate decision therapy decision algorithms run in real time. As such, the fifth research question became: Do the retrospective strategies work for real time decision making?

Chapter 2 - Methods

2.1 – Data Description

55 Cardiologist adjudicated manufacturer case reports with at least one therapy delivery event were provided in the form of pdf files by Dr. Selçuk Adabag in accordance with an institutional review board (IRB) approved protocol and stored in accordance with this protocol. Table 1 below contains information on all 55 of the received case reports. Upon receiving the case reports, the EGM signals of each case report were reviewed for their ability to be digitized into ASCII. 5 case reports were excluded at this point because they had thick solid lined grids obscuring much of the EGM signals on the pdf. 1 other case report was excluded because it came from a different manufacturer than all of the other devices and did not provide at least 20 RR intervals of EGM signal prior to therapy required for analysis. This left 49 acceptable case reports with a total of 54 therapy events.

Table 1: Overall information of received case reports. Excluded case reports have the "PDF to ASCII transfer (yes/no) column filled with red."

Case	# of Appropriate therapy events	# of Inappropriate therapy events	Therapy rhythm	PDF to ASCII transfer (yes/no)	Device Type	Manufacturer	# of RR intervals prior to therapy available for digitization up to 40
1		**1**	AF w/ RVR	yes	Single Chamber	Medtronic	40
2	**2**			no	Single Chamber	Medtronic	26 & 26
3	**1**		VT	yes	Single Chamber	Medtronic	40
4	**4**		VF	no	Dual Chamber	Biotronik	< 20
5	**1**		VT	yes	Single Chamber	Boston Sci	30
6	**1**		VT	yes	Single Chamber	Boston Sci	30
7	**1**		VF	yes	Subcutaneous	Boston Sci	40
8	**1**		VF	yes	Subcutaneous	Boston Sci	40
9	**1**		VF	yes	Subcutaneous	Boston Sci	40
10	**1**		VF	no	Single Chamber	Medtronic	25
11	**1**		VT	yes	Single Chamber	Boston Sci	29

12	**1**		VT	yes	Single Chamber	Boston Sci	30
13	**1**		VT	yes	Single Chamber	Boston Sci	29
14	**1**		VT	yes	Single Chamber	Boston Sci	29
15	**1**		VT	yes	Single Chamber	Boston Sci	29
16	**1**		VT	yes	Single Chamber	Boston Sci	29
17	**1**		VT	yes	Single Chamber	Boston Sci	30
18	**2**		VT	yes	Single Chamber	Boston Sci	30 & 30
19	**1**		VT	yes	Single Chamber	Medtronic	26
20	**1**		VT	yes	Single Chamber	Medtronic	26
21	**1**		VT	yes	Single Chamber	Medtronic	40
22		**1**	AF	yes	Single Chamber	Boston Sci	40
23		**2**	AF	yes	Single Chamber	Boston Sci	40 & 34
24		**1**	AF	yes	Single Chamber	Boston Sci	40
25		**1**	AF	yes	Single Chamber	Boston Sci	40
26		**1**	AF	yes	Single Chamber	Boston Sci	40
27		**2**	AF	yes	Single Chamber	Boston Sci	40 & 27
28	**1**		VT	yes	Single Chamber	Medtronic	40
29	**1**		VT	yes	Single Chamber	Medtronic	40
30	**1**		VT	yes	Single Chamber	Medtronic	26
31	**1**		VT	yes	Single Chamber	Medtronic	25
32	**1**		VT	yes	Single Chamber	Medtronic	25
33	**1**		VF	no	Dual Chamber	Medtronic	40
34	**1**		VF	yes	Single Chamber	Medtronic	26
35	**1**		VT	yes	Subcutaneous	Boston Sci	40
36	**1**		VF	yes	Subcutaneous	Boston Sci	40
37	**1**		VF	yes	Subcutaneous	Boston Sci	40
38	**1**		VF	yes	Subcutaneous	Boston Sci	40

39	**1**		VF	yes	Subcutaneous	Boston Sci	40
40	**1**		VF	yes	Subcutaneous	Boston Sci	40
41	**1**		VF	yes	Subcutaneous	Boston Sci	40
42	**1**		VF	yes	Subcutaneous	Boston Sci	40
43	**1**		VT	yes	Subcutaneous	Boston Sci	40
44	**1**		VF	yes	Subcutaneous	Boston Sci	40
45	**1**		VF	yes	Subcutaneous	Boston Sci	40
46	2		VT	yes	Single Chamber	Medtronic	28 & 25
47		**2**	AF	yes	Single Chamber	Medtronic	40 & 21
48	**1**		VT	yes	Single Chamber	Medtronic	26
49	**1**		VT	yes	Subcutaneous	Boston Sci	40
50		**1**	NSVT	yes	Subcutaneous	Boston Sci	40
51	**1**		VT	yes	Subcutaneous	Boston Sci	40
52	**1**		VT	yes	Single Chamber	Boston Sci	33
53	**1**		VT	yes	Single Chamber	Boston Sci	33
54	**1**		VF	no	Single Chamber	Medtronic	40
55	**1**		VF	no	Single Chamber	Medtronic	40

All 49 acceptable case reports were from either single chamber TV-ICD or S-ICD devices, two of the most commonly implanted ICD types. A single chamber TV-ICD is an ICD with a single transvenous lead implanted directly into the right ventricle (see Figure 3). When therapy is delivered, it is done so through this lead. Dual chamber and Biventricular TV devices are less common and are only indicated when pacing of both the atria and ventricles is indicated. Rather than transvenous leads, the S-ICD has a single lead placed directly over the right ventricle just above the rib cage (see Figure 3). By placing the lead subcutaneous and not transvenous, the S-ICD overcomes a variety of issues with traditional TV-ICDs including lead fracture, cardiac perforation, etc. However, they still deliver inappropriate therapy at similar rates as TV-ICDs [8], but typically due to oversensing and not SVTs.

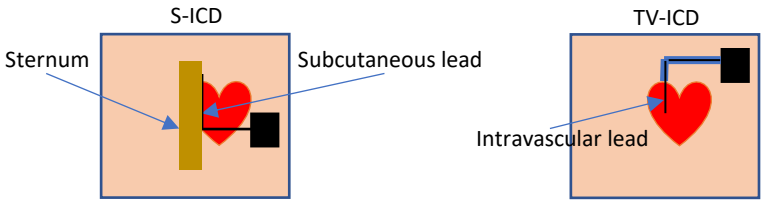


Figure 3: Picture of implanted S-ICD (left) and single chamber TV-ICD (right), box represents patient's chest. Note that the TV-ICD has a single lead implanted directly into the heart via the venous system while the S-ICD system has a single lead implanted subcutaneously just to the left (from the patient's perspective) of the sternum.

The 49 acceptable case reports with 54 total therapy events had their far-field or "Shock" EGM digitized for up to 40 RR intervals prior to delivery of therapy for each event with at least 20 RR intervals of non-paced or shocked EGM on the case report. All case reports from S-ICD devices had at least 40 RR intervals of EGM prior to the therapy, but a majority of the single chamber TV-ICD devices did not and thus the total EGM signal length varies from 21 RR to 40 RR intervals prior to therapy delivery. See acceptable forms of sample data in Figure 4 below. Digitization was completed with MATLAB 2020a (The MathWorks, Inc.) via a custom written MATLAB script. Following digitization, all signals were visually adjudicated to ensure all EGM signals were transferred into ASCII with high fidelity. For a detailed description of the digitization process see Appendix A.1.

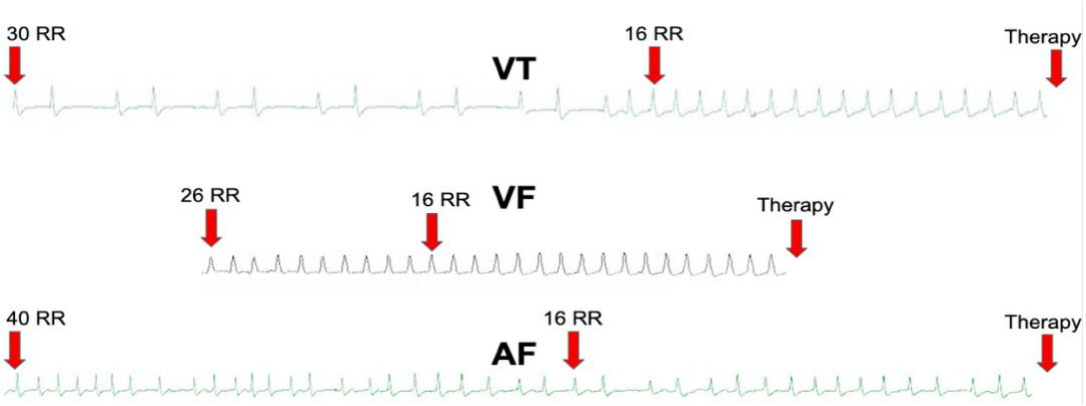


Figure 4: Representative examples of acceptable far-field or "Shock" EGM signal for digitization. **Top)** signal adjudicated as VT, **middle)** signal adjudicated as VF and **bottom**) signal adjudicated as AF. Red arrows indicate important events such as the start of the signal with the number of RR intervals present, 16 RR intervals prior to the therapy and therapy delivery (not shown in the signal). These are cropped and modified pdf images taken from the case reports that are then stitched together. For a detailed description of the digitization process see Appendix A.1.

42 of the 54 the digitized events were deemed appropriate based upon adjudication. 12 were labeled inappropriate. AF accounted for 11 of the inappropriate events, all occurring in single chamber TV devices, and there was 1 event of NSVT occurring with a S-ICD device. The dataset is made up of three device classes: Visia AF MRI VR / Evera XT VR (Medtronic) single chamber TV-ICDs, Dynagen EL ICD D150 (BSci) single chamber TV-ICDs and Emblem MRI S-ICD A219 (BSci). The dataset is summarized in Table 2 below.

Table 2: Summary of the dataset of digitized events. The parenthesis indicates the number of shocks delivered as therapy out of the total therapies delivered in each category.

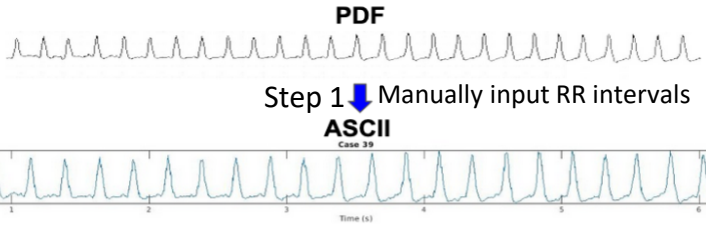
	Appropriate therapy (n=42)		**Inappropriate therapy (n=12)**	
	VT	VF	NSVT	AF

Visia AF MRI VR / Evera XT VR (Medtronic)	12 (1)	1 (0)	0	3 (1)
Dynagen EL ICD D150 - VR (BSci)	13 (1)	0	0	8 (2)
Emblem MRI S-ICD A219 (BSci)	4 (4)	12 (12)	1 (1)	0

RR intervals as detected by the device were given on each case report. BSci single chamber TV-ICD case reports gave RR intervals to the exact ms and Medtronic single chamber TV-ICD case reports gave RR intervals rounded to 10 ms. S-ICD case reports did not give values for RR intervals, but still marked them on the document in a normal ECG grid (200 ms full thickness lines and 40 ms dotted lines). Thus, they could be estimated to the nearest 40 ms. After manually estimating the RR intervals for the 17 S-ICD events, the RR intervals for each event were manually transferred into ASCII format from the case report.

2.2 – Data Analysis Workflow

All data was analyzed via the workflow presented in Figure 5 on the next page. First, in step 1, pdf images of the EGM signals were digitized into ASCII as described in Appendix A.1. In step 2, the ASCII EGM signals and RR interval data for each event were segmented into the correct window size and 11 metrics, 4 novel nonlinear-based and 7 RR-based were calculated. The 4 nonlinear-based metrics required the EGM signals and 7 RR-based required the RR interval data. Once all metrics were calculated for a dataset, step 3 was completed, in which the metrics were normalized via a Z-score using the mean and standard deviation of the dataset and input into a PCA or linear combination score for discrimination.



Nonlinear-based analysis				Standard RR-based analysis						
MSF	MSE	SE	Kurtosis	MeanRR	pNN50	RMS	STD	SD1	SD2	SD1/SD2
4.362	0.736	8.218	4.087	248.846	0.000	10.583	9.519	7.616	11.101	0.686

Figure 5: Workflow for data analysis. **Step 1)** PDF image data was converted into ASCII and RR intervals manually entered. **Step 2)** The EGM signals and RR intervals were windowed **(2.1)** and then the Nonlinear-based metrics were calculated using the ASCII EGM signal **(2.2)** and Standard RR-based metrics calculated with the manually input RR intervals **(2.3)**. **Step 3)** The metrics were then normalized with a Z-score and input into a PCA or therapy score system or analyzed with a boxplot.

2.3 – Step 2.2: Nonlinear-based metrics

4 novel Nonlinear-based metrics were utilized in this study: Multi-Scale Entropy (MSE), Shannon Entropy (SE), Multi-Scale Frequency (MSF) and Kurtosis. They were calculated using the windowed ASCII far-field/"Shock" EGM signals from each event. A detailed description of the techniques is available in Appendix A.2. MSE & SE should capture the randomness of the AF signals compared with the VT/VF signals, while AF should have a different MSF due to its erratic patterns. Moreover, Kurtosis should be different for the AF signals due to their different morphology when compared to VT/VF signals. Thus, these 4 novel techniques should be significantly different between the inappropriate (AF) and appropriate (VT/VF) events.

2.4 – Step 2.3: Standard RR-based metrics

7 RR-based metrics were utilized in this study: the mean RR interval (MeanRR), proportion of the successive differences in RR intervals (NNs) that exceeds 50 ms (pNN50), root mean square of the NNs (RMS), standard deviation of the NNs (STD), and

the three Poincare metrics (SD1, SD2 and SD1/SD2). They were calculated using the manually input RR intervals from each event. A detailed description of the techniques is available in Appendix A.3. These techniques are similar to currently utilized stability measures and thus should show a difference between the inappropriate (AF) and appropriate signals (VT/VF) as AF typically has less stable RR intervals.

2.5 – Step 3: Principal Component Analysis

Principal component analysis (PCA) is commonly utilized to shrink the dimensionality of data. In this study, we have 11 metrics for each event. Making sense of the 11 metrics is challenging, but it is much easier to understand if we have 2 or 3 metrics to separate the data. Thus, PCA was utilized to reduce the dimensionality of the data to 2 or 3 dimensions rather than 11.

PCA transforms the large data set into a smaller one by creating linear combinations of the larger data set with the most possible variance. The most common approach when using PCA is to choose the number of principal components (PCs) to represent the data with based on the number of PCs that explain at least 95% of the variability in the data. However, given the desire to have the data be explained in only 2-3 metrics, the PCs are limited to 2-3 in this study, even though they may not contain 95% of the variability. Thus, the results of PCA on a given dataset are presented as the first 2-3 PC scores. PC coefficients greater than 0.3 are considered significant in this study [12].

2.6 – Step 3: Boxplots and Statistics

Boxplots of the normalized data were created to analyze the different studies with MATLAB. All boxplots were analyzed with a one-way ANOVA with statistical significance for $p<0.05$.

Chapter 3 – Results

3.1 – Research Question 1 – Does a long or short RR window prior to therapy lead to better retrospective discrimination?

The first research question being asked was how many RR intervals one should consider prior to therapy when retrospectively evaluating therapy events. This question was being asked to determine if using a longer window led to more significantly different features than a shorter window just prior to therapy. Thus, the first study completed was a retrospective analysis looking at all RRs available prior to the therapy vs 16 RRs prior to therapy. When all RRs were used there was an uneven number of RR intervals between EGM signals, ranging from 40 RRs to 21 RRs in length. The EGM signals often began in sinus rhythm and transitioned into arrythmia at some point. In some events, the EGM signal was entirely arrythmia prior to therapy. With the 16 RR window, each signal was now a uniform 16 RRs in length. 16 RRs was chosen because upon visual inspection, this was the number of RRs for which every appropriate event EGM signal had progressed into VT/VF i.e., the entire signal is VT/VF for every appropriate event.

Following windowing, box plots were created for the 11 metrics to compare the appropriate and inappropriate events. Only MSF and pNN50 were significantly different between the appropriate and inappropriate events for the all RRs window (Figure 6). Shrinking the window to be arrythmia only for the appropriate events (the 16 RR window), a majority of parameters became significantly different between the appropriate and inappropriate events, except for SD1/SD2 and MSE (Figure 7).

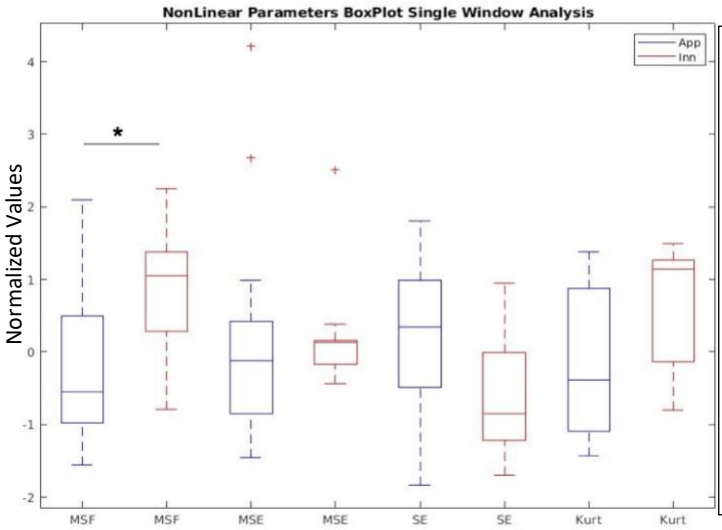
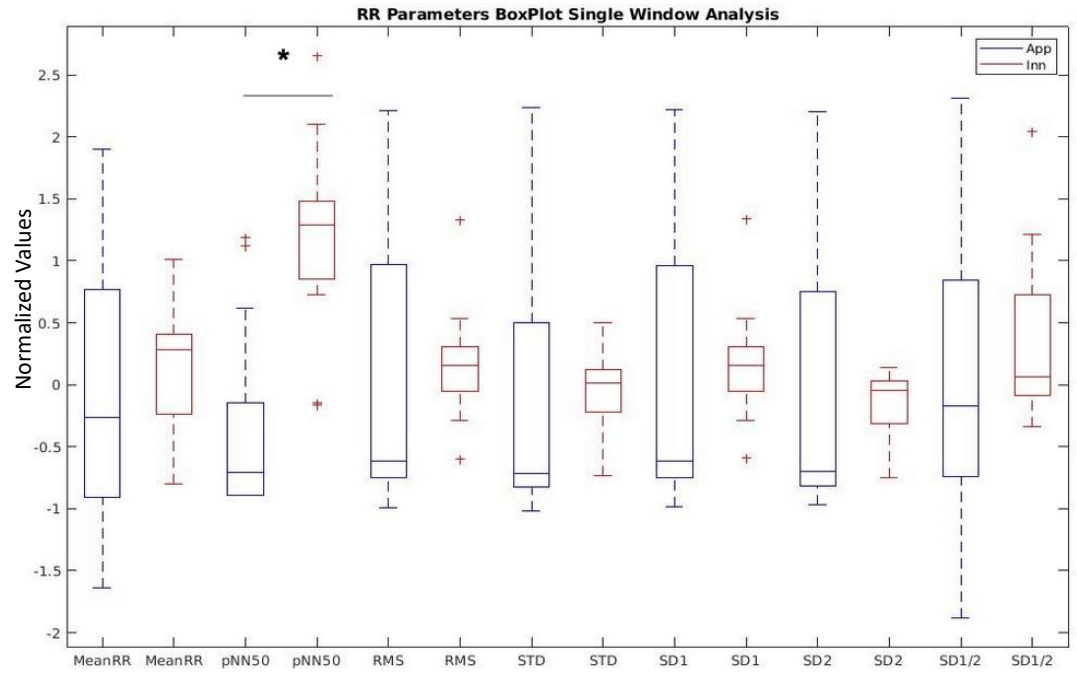


Figure 6: Boxplots of the nonlinear-based (top) and RR-based (bottom) metrics across all RRs. * given for statistical significance between groups of appropriate and inappropriate.



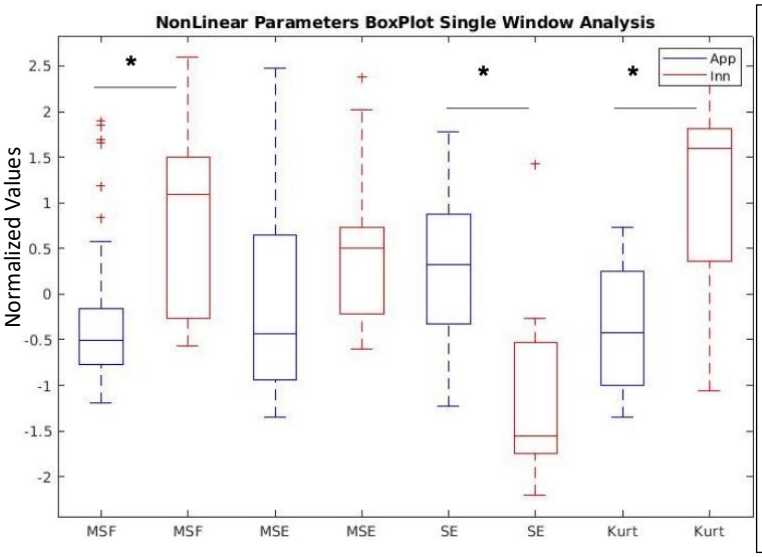
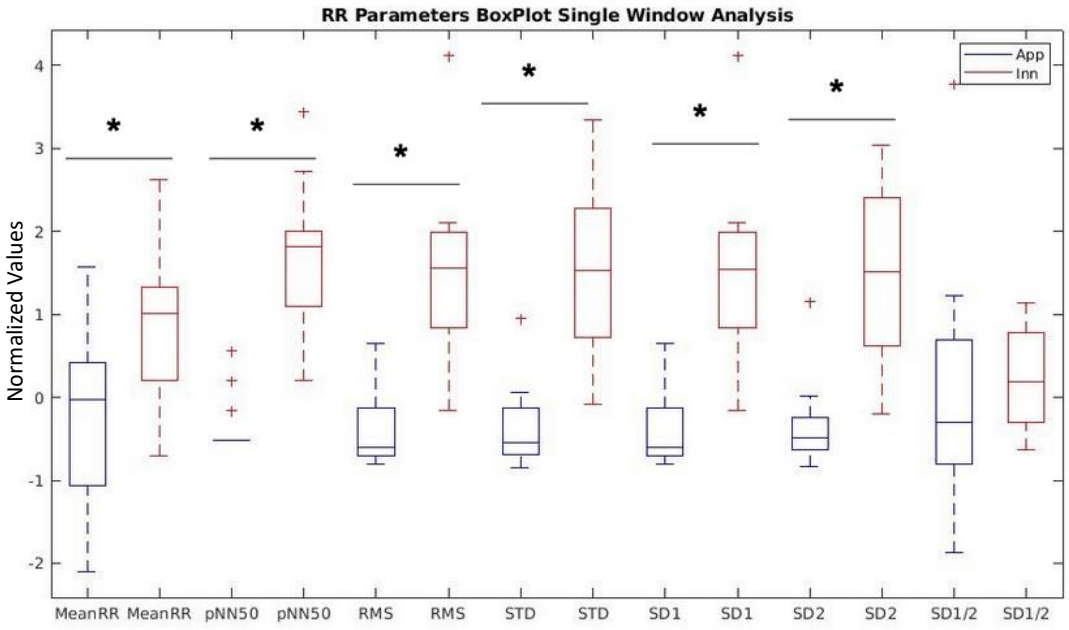


Figure 7: Boxplots of the nonlinear-based (top) and RR-based (bottom) metrics for the 16 RRs prior to therapy delivery. * given for statistical significance between groups of Appropriate and Inappropriate.



The boxplot analysis was followed up by completing PCA on both datasets. The results of the PCA with the first two PC scores are given below in Figures 8 & 9. No clear visual discrimination can be made between the appropriate and inappropriate events when using the features of the all RRs dataset (Figure 8). However, the shorter 16 RR window dataset (Figure 9) showed clear visual separation between the appropriate and inappropriate events.

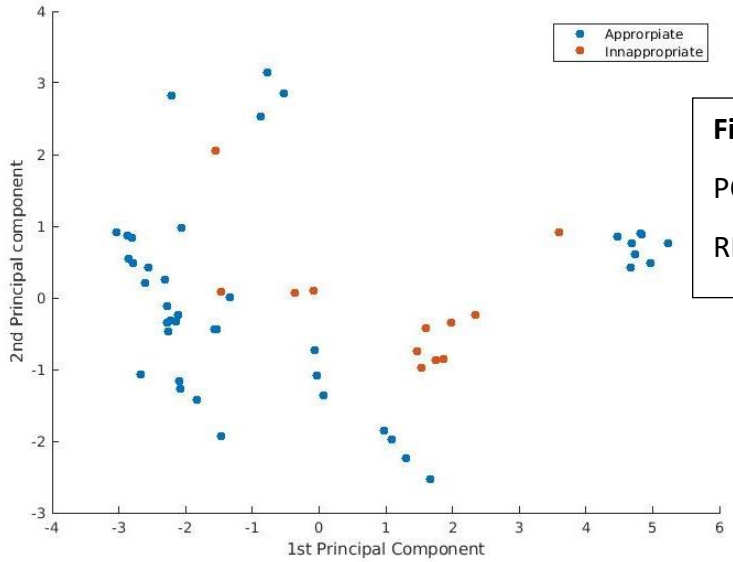


Figure 8: First two PC scores for the all RR window.

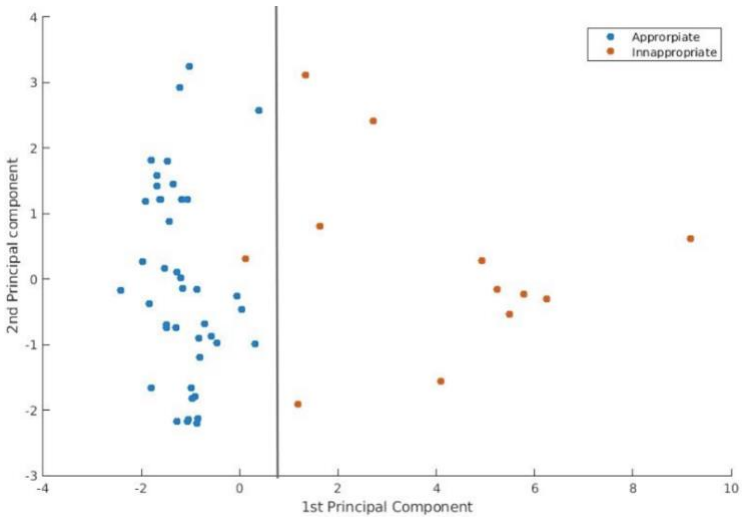


Figure 9: First two PC scores for the 16 RR window prior to therapy delivery. A solid black line was placed to aid in visual discrimination.

A 3-D plot of the first three PCs was also desired to see if better visual discrimination was possible by adding in the third PC. This was only done with the 16 RR window dataset. Similar levels of visual discrimination are possible on this plot, Figure 10 below, when compared to the 2-D plot of just the first two PCs above in Figure 9.

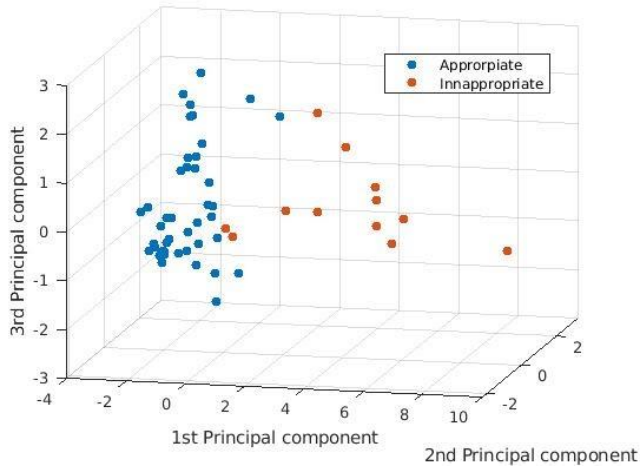


Figure 10: First three PC scores for the 16 RR window prior to therapy delivery.

Overall, the answer to this research question is that the 16 RR window is superior to the all RR window when retrospectively considering if therapy is appropriate or inappropriate. This is most likely because the NSR portions in some of the appropriate events of the all RR windows was tricking the metrics and PCA into thinking the rhythm was more NSR like than VT/VF like. Thus, in order for the metrics plus PCA to accurately separate appropriate therapy from inappropriate therapy, the window size must be chosen such that all of the appropriate events are in VT/VF for the entire duration of the window. In this study, this is the 16 RR window.

3.2 – Research Question 2 – Does arrythmia type and/or manufacturer influence discrimination?

Following the success of the 16 RR window dataset with PCA to discriminate between the appropriate and inappropriate events, it was desired to see if the different arrythmias (VT, VF, AF, NSVT) could be separated from each other as well. As such, the events were sorted by arrythmia type – VT, VF and AF – for the below boxplot analysis in Figure 11. AF accounted for 11 of the 12 inappropriate events while VT/VF accounted for all 42 of the appropriate events. The singular NSVT event was removed because there was only one of them. We see that between VT or VF and AF the same parameters remain statistically significant as in Figure 7 above.

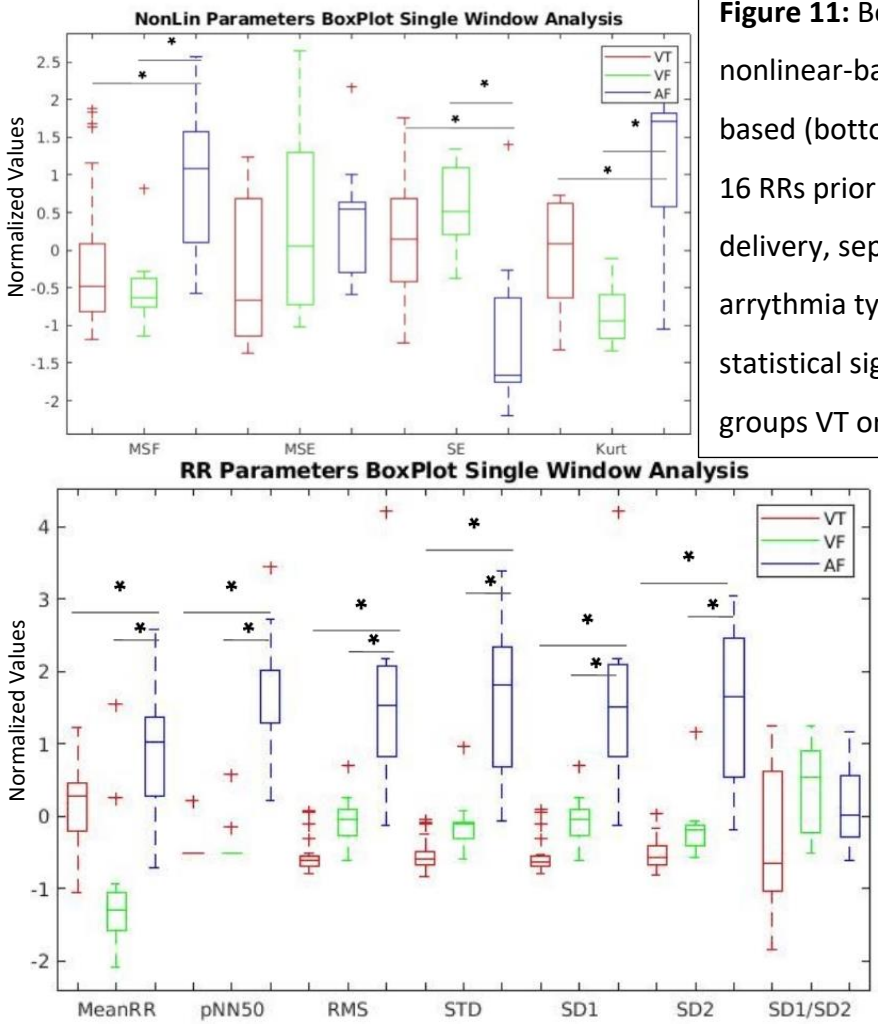


Figure 11: Boxplots of the nonlinear-based (top) and RR-based (bottom) metrics for the 16 RRs prior to therapy delivery, separated by arrythmia type. * given for statistical significance between groups VT or VF and AF.

Next, PCA was completed on the 16 RR dataset once again. The PCA is identical to the one shown with the 16 RR dataset (Figure 9) except for how the resulting 2-D (first two PCs) and 3-D (first three PCs) plots are labeled. In these plots, Figures 12 and 13 below, the events are labeled by their arrythmia type VT or VF for appropriate events and AF or NSVT for the inappropriate events. The inappropriate events are also labeled by manufacturer, BSci or Medtronic. With this analysis, it was seen that the VT and VF groups were not clearly visually distinguishable from one another but remained distinguishable from the inappropriate therapy rhythm groups. While VT and VF signals typically have different morphology, they do tend to have similar regularity to their RR intervals. As such, the aforementioned result was not surprising. Figures 12 & 13 also show that most of the inappropriate BSci events are further separated from the appropriate events when using either 2 or 3 PCs compared to the inappropriate Medtronic events.

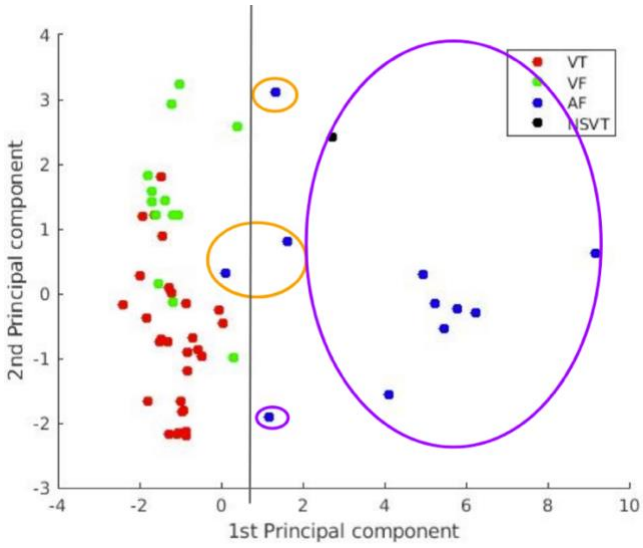


Figure 12: First two PC scores for the 16 RR window prior to therapy delivery but separated by arrythmia type and manufacturer. Inappropriate BSci cases surrounded by purple ovals while inappropriate Medtronic cases are surrounded by orange ovals. A solid black line was placed to aid in visual discrimination.

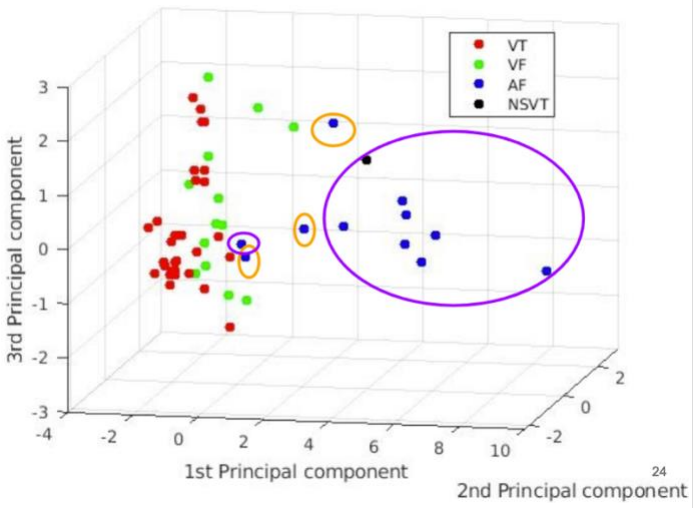


Figure 13: First three PC scores for the 16 RR window prior to therapy delivery but separated by arrythmia type and manufacturer. Inappropriate BSci cases surrounded by purple ovals while Inappropriate Medtronic cases are surrounded by orange ovals.

Digging further into the PCA analysis of the 16 RR window, the loading coefficients for the first three PCs were analyzed to see which parameters were contributing a significant portion to the formation of the PC score. These are given below in Table 3. Across the first three PCs every parameter contributed in a significant way to at least one PC. The first PC was primarily composed of the RR-based metrics while the second PC was more heavily biased by the nonlinear-based metrics. PC3 was heavily influenced by MSF, MSE and SD1/SD2.

Table 3: Loading coefficients for the PCA of Study 2 and Study 3. Significant contributions are bolded.

	Loading Coefficients		
	PC1	PC2	PC3
MSF	0.272	-0.148	**0.438**
MSE	0.068	**0.402**	**0.705**
SE	-0.285	**0.336**	0.064
Kurt	**0.303**	**-0.400**	0.042
MeanRR	0.244	**-0.416**	-0.162
pNN50	**0.375**	0.118	-0.006
RMS	**0.367**	0.219	-0.122
STD	**0.376**	0.174	-0.050
SD1	**0.367**	0.220	-0.124

SD2	**0.370**	0.131	0.011
SD1/2	0.027	**0.461**	**-0.497**

Lastly, it was desired to see how many PCs actually encompassed 95% of the variability with the PCA of the 16 RR window. Thus, Figure 14 below was created which shows the percent variance explained by each PC up to 95% of the total. First, it was seen that it requires 5 PCs to cover 95% of the total variance in the data. However, the first two PCs represent roughly 80% of the variance, and the first three represent almost 90% of the variance. Thus, while not meeting the typical requirement of 95% of variability covered in the utilized PCs, the first two and three PCs do cover a majority of the variability in the data.

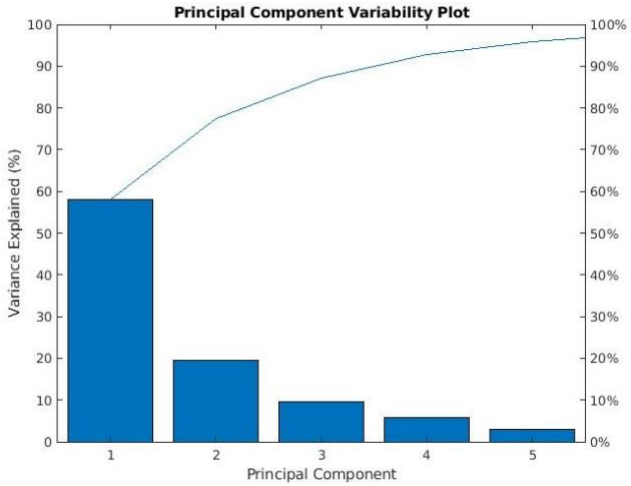


Figure 14: Bar plot showing the percent variance explained by each PC up to 95% of the total variance. The cumulative total variance explained as the amount is summed from PC 1 to 5 is given by the solid line.

Overall, answering this research questions showed that the inappropriate BSci events were typically further separated from the appropriate events compared to the inappropriate Medtronic events. This may be a result of the programming of the devices. All 3 Medtronic events had stability analysis turned on at either 40 or 50 ms, while most of the BSci events had it turned off. Stability analysis (described in the introduction) is similar to the RR-based metrics used in this study. As such, if turned on, it should have prevented inappropriate therapy for highly variable RRs. PC1 is primarily composed of the RR-based metrics, and thus a PC1 greater than a certain value is computing a similar decision to stability analysis. Thus, most likely, most of the events of inappropriate

therapy on BSci devices would have been prevented by programming the stability analysis on.

This finding begs an important question: what is the advantage of these metrics plus PCA when compared with existing algorithms to retrospectively separate the signals? The advantage is that the combination of the seven RR-based metrics should provide better sensitivity than a single metric as used in the stability analysis. This is seen in that 2 out of the 3 Medtronic events lie to the right of the black line in Figure 12, even though they had stability analysis turned on.

3.3 – Research Question 3 – Does separating the data by manufacturer into two separate datasets improve discrimination?

Given the findings of research question 2 that the inappropriate Medtronic events were less separated from the appropriate events compared to most of the BSci events, the two manufacturers' devices were individually investigated. It was hoped that separating them would lead to better visual separation between the appropriate and inappropriate events. To this end, the 16 RR dataset was separated into a Medtronic dataset and BSci dataset after all metrics were calculated and before normalization. Normalization with a Z-score was then completed and PCA completed on each dataset. Figures 15 and 16 below show the results of the PCA. In both figures, all inappropriate events (AF) are clearly separated from the appropriate events (VT/VF) based on visual discrimination.

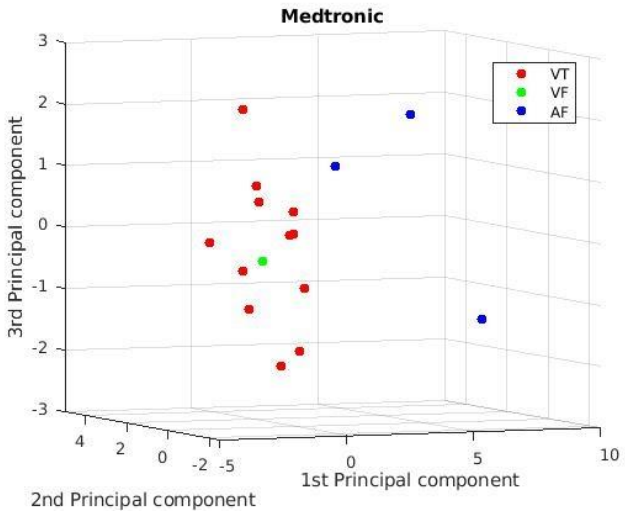


Figure 15: First three PC scores for the 16 RR window prior to therapy delivery, but only for Medtronic single chamber TV devices.

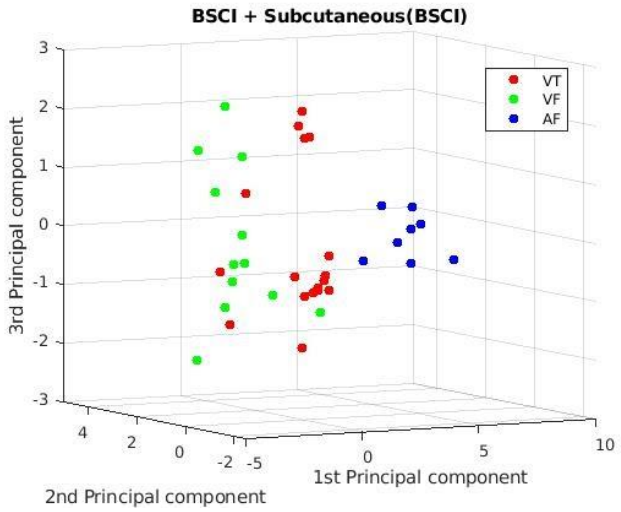


Figure 16: First three PC scores for the 16 RR window prior to therapy delivery, but only for BSci single chamber TV devices and BSci S-ICD devices.

To investigate why there was better separation when doing PCA on the two separated data sets, the loading coefficients for the first three PCs were compiled in Tables 4 & 5 below. The loading coefficients in Table 4 for the BSci devices remained similar to those for the entire dataset in Table 3. However, the loading coefficients for the separated Medtronic dataset in Table 5 had significant change. The MeanRR, STD and SD2 did not significantly contribute to any of the first three PCs, MSE contributed much more to PC1 and not the second two, MSF contributed much more to PC2 rather than PC3, SE contributed more to PC2 and PC3 and less to PC1 and lastly Kurtosis switched its significant contributions from PC1 and PC2 to PC3.

Table 4: Loading coefficients for the PCA of the BSci dataset.

	Loading Coefficients		
	PC1	PC2	PC3
MSF	0.268	-0.182	**0.395**
MSE	0.043	**0.344**	**0.731**
SE	**-0.305**	**0.334**	0.158
Kurt	**0.315**	**-0.377**	0.006
MeanRR	**0.281**	**-0.367**	-0.147
pNN50	**0.377**	0.106	0.000
RMS	**0.354**	0.261	-0.089
STD	**0.365**	0.211	-0.013
SD1	**0.354**	0.262	-0.091

SD2	**0.359**	0.160	0.050
SD1/2	-0.002	**0.492**	**-0.494**

Table 5: Loading coefficients for the PCA of the Medtronic dataset.

	Loading Coefficients		
	PC1	PC2	PC3
MSF	0.153	**-0.374**	0.123
MSE	**0.497**	-0.174	-0.022
SE	-0.061	**0.616**	**-0.481**
Kurt	-0.047	-0.124	**0.552**
MeanRR	-0.179	0.269	0.219
pNN50	**0.437**	0.283	0.209
RMS	**0.316**	0.197	0.136
STD	0.287	0.204	0.122
SD1	**0.317**	0.197	0.137
SD2	0.255	0.205	0.105
SD1/2	**0.390**	**-0.348**	**-0.541**

Overall, splitting up the data into two datasets, one for each manufacturer, led to good visual separation with both datasets. This is because the PCA relied upon a unique combination of measures for creating the first few PCs for the Medtronic dataset. In doing so, it visually separated all 3 inappropriate Medtronic events from the appropriate Medtronic events even though they had regular RRs and thus stability analysis failed to withhold therapy. Despite these promising results, separating the datasets is not ideal as this would mean for all future investigations the datasets would need to be split based upon manufacturer.

3.4 – Research Question 4 – Do the therapy scores discriminate appropriate and inappropriate therapy?

The last retrospective study completed was to look at using a linear combination score termed “therapy scores” to discriminate between the appropriate and inappropriate events. The therapy scores were separated into a nonlinear therapy score (NonLinScore)

and RR therapy score (RRScore). The NonLinScore was formed as NonLinScore = MSF – SE + Kt. The RRScore was formed as RRScore = MeanRR + pNN50 + RMS + STD + SD1 + SD2. Both scores were formed by simply summing the statistically significant parameters from the boxplot analysis in study 2 (Figure 9). The only exception to summing was a subtraction of SE because SE was significantly decreased in the Inappropriate events when compared with the Appropriate events in Figure 9. The insignificant metrics MSE and SD1/SD2 were left out the therapy scores.

The innovative idea for the therapy scores came about because when looking at the previous PCA analysis in research questions 1-3 it was seen that parameters such as SD1/SD2 and MSE were making significant contributions to the PCs even though they were not statistically different between the inappropriate and appropriate groups. This is because PCA shrinks the dimensionality by looking at how much variability across the entire dataset can be explained by each metric. So even though they weren't statistically different, there was still a large amount of variability in the entire dataset with these metrics and thus they were significantly utilized in the creation of the first few PCs. The therapy scores also make it clear to see the difference in the RR-based metrics and the Nonlinear-based metrics between the appropriate and inappropriate events.

Figure 17 below shows the results of using the RRScore and NonLinScore to separate the 16RR dataset rather than PCA. Similar visual discrimination was achieved with this method compared to PCA. However, rather than representing the discrimination with a line, a black box of RRScore and NonLinScore < 2.5 is utilized. This region is thusly termed the "Appropriate Therapy Zone" or ATZ. All appropriate events are within the ATZ while only 1 inappropriate event is inside it. Moreover, the utility of the ATZ can be seen in that 3 Inn events were only visually distinguishable based upon their RRScore (RRScore > 2.5 & NonLinScore < 2.5) while 1 inappropriate event was only distinguishable upon its NonLinScore (RRScore < 2.5 & NonLinScore > 2.5). Thus, if only the RRScore or only the NonLinScore was used to determine if the therapy was appropriate rather than the ATZ, the method would be less accurate.

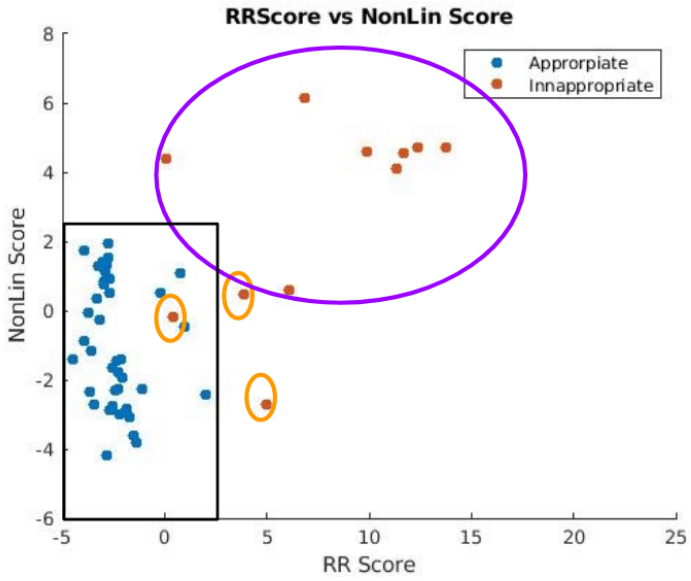


Figure 17: RRScore on the x-axis and NonLinScore on the y-axis to visually discriminate between Appropriate and Inappropriate cases. The ATZ is the area

3.5 – Research Question 5 – Do the retrospective strategies work for real time decision making?

Research questions 1-4 were all retrospective studies looking at a window of RRs from the known therapy backwards. Thus, with research question 5, it was desired to investigate if PCA or the therapy scores and ATZ could visually distinguish whether or not to provide therapy across time. In this sense, once a datapoint moved into a specified area the device would deliver therapy.

5 events (2 Appropriate and 3 Inappropriate) directly following a previous therapy were removed from the dataset as the focus of this study was the decision to deliver the first therapy. Next, 25 RR interval window lengths of each EGM signal were then obtained. Each 25 RR interval window was then split into 4 successive windows of 10 RR, each with 5 RR overlap. See Figure 18 below to visualize the sliding windows. Each 10 RR window was then normalized with the Z-score using the mean and standard deviation across all of the 10 RR windows. Boxplots for the first and fourth windows were then generated to compare the measures across time Figures 19 and 20. This "pseudo" real time analysis is thusly looking at a short window a few seconds prior to the delivery of the first therapy and then at another short window right before the first therapy is delivered. With window 1 the only statistically significant parameter was MSF, while with window 4, all parameters except for MSE, MeanRR and SD1/SD2 were

statistically significant. This illustrates a change in the properties of many of the appropriate EGM signals from window 1 to window 4. This is expected because many of the appropriate EGM signals transition from a NSR to VT/VF between window 1 and 4.

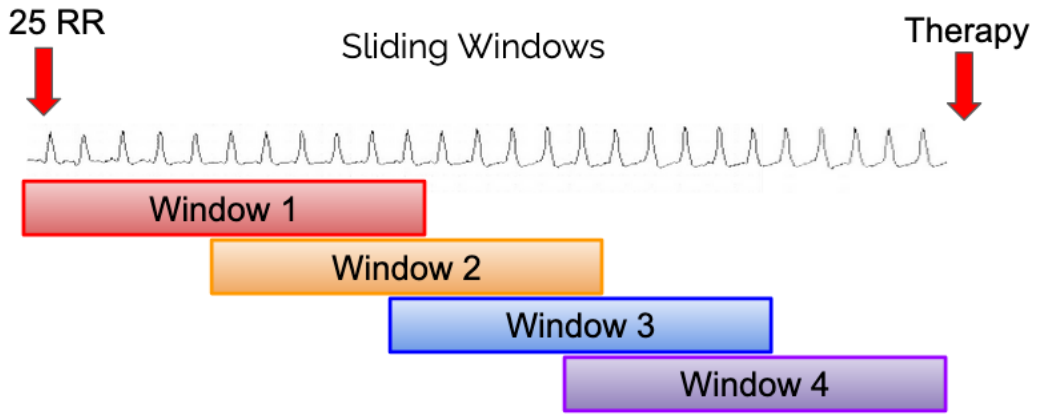


Figure 18 : 4 windows of 10 RR length created from 25 RR prior to therapy.

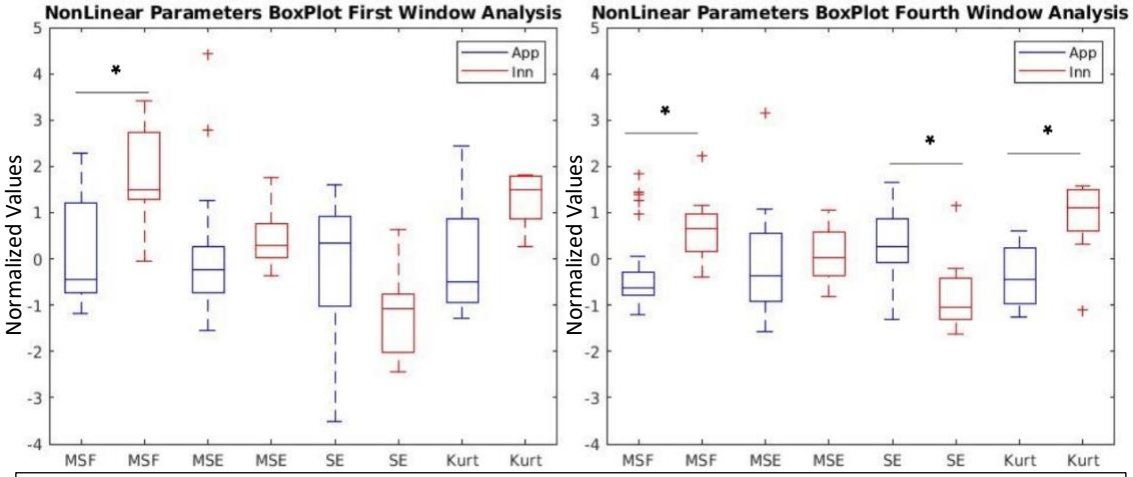


Figure 19: Boxplots of the Nonlinear-based metrics for the first (left) and fourth (right) windows comparing the Appropriate and Inappropriate cases.

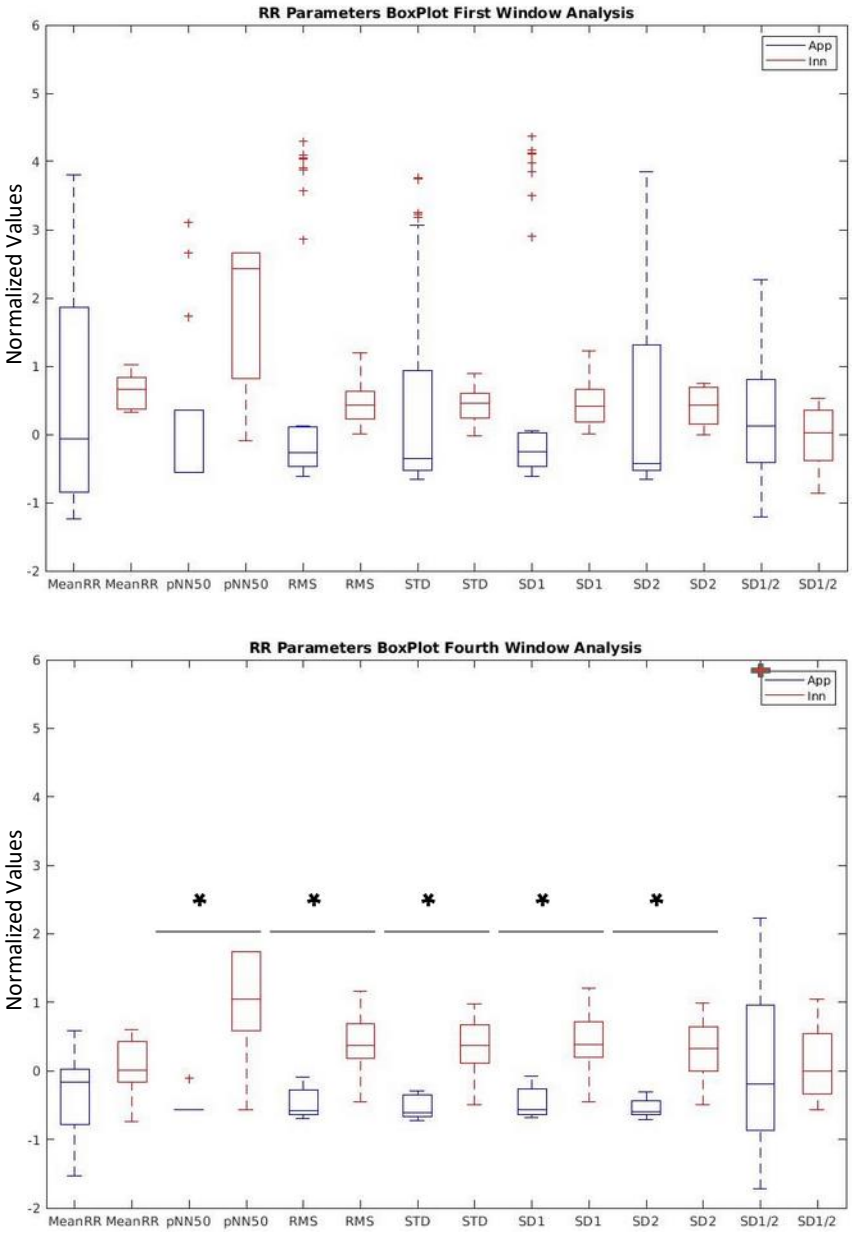


Figure 20: Boxplots of the RR-based metrics for the first (top) and fourth (bottom) windows comparing the Appropriate and Inappropriate cases.

PCA was then completed with all of the 10 RR windows to generate the 1st and 2nd PC for each window (Figure 21). With window 1, we see no clear visual separation between the inappropriate and appropriate events. However, as we move from window 1 and window 2 to window 3, we see movement of all of the appropriate events to a PC1 of less than 0, while most inappropriate events maintain a PC1 greater than 0. The hope was then that all of the appropriate events would move into the specified area (PC1 < 0) by window 4 and none of the inappropriate events to do so. However, 2 inappropriate events moved into the specified area by window 4. Moreover, it is important to note that a

majority of the appropriate events began in the specified area in Window 1. This is because a majority of the events began in VF/VT. There is then limited movement of these events over time because their rhythm remains VT/VF for all 4 windows.

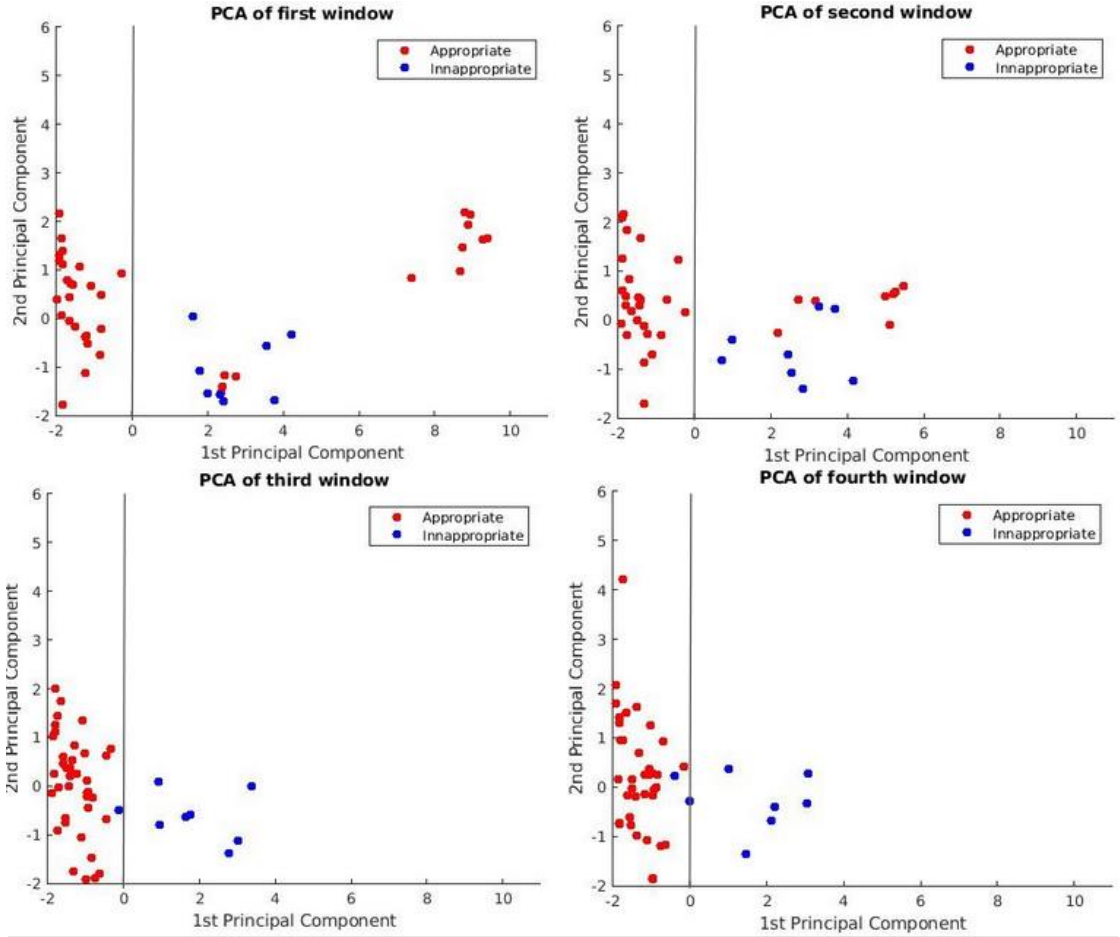


Figure 21: The first two PC scores for each 10 RR window dataset: Window 1 (top left), Window 2 (top right), Window 3 (bottom right) and Window 4 (bottom left). Black line added to aid with visual discrimination.

Beyond PCA, the use of the therapy scores was also investigated for its ability to separate the signals in a "pseudo" real time fashion. As such, instead of completing PCA, the previously described RRScore and NonLinScore were applied to the datasets of windows 1-4. The results of this investigation are given below in Figure 22. Based on the visual separation in Figure 22, the ATZ was determined as the region of a RRScore less than 0 and NonLinScore less than 2.5. In windows 1 and 2, a slight majority of the number of appropriate events are within the ATZ while none of the inappropriate events. By

window 3 all but 1 appropriate event has entered the ATZ, which subsequently enters the ATZ in window 4. Meanwhile, none of the inappropriate events ever enter the ATZ. This is more ideal when compared with the PCA.

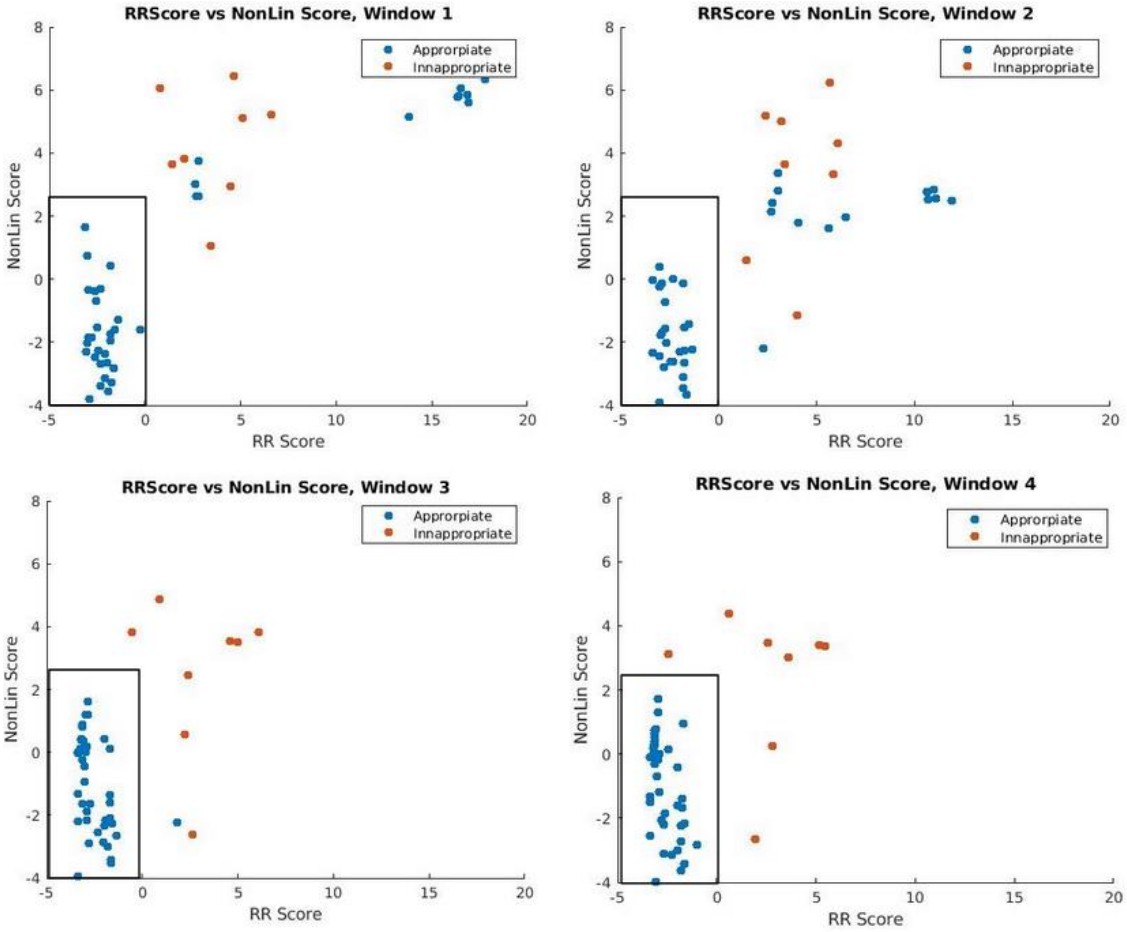


Figure 22: Plots of the RRScore vs. NonLinScore for each 10 RR window dataset: Window 1 (top left), Window 2 (top right), Window 3 (bottom right) and Window 4 (bottom left). The ATZ is given by the black box.

Chapter 4 – Summary

4.1 – Major Conclusions

The primary goal of this thesis was to develop novel strategies for discrimination between inappropriate and appropriate ICD therapy. The novelty of the new strategies is that they utilize a combination of 4 novel nonlinear metrics in addition to 7 standard RR-based metrics fed into PCA or linear combination scores (therapy scores) to discriminate rather than traditional rate, morphology, stability and onset analysis to determine if a rhythm is appropriate for therapy. In evaluating the novel strategies, the thesis addressed 5 primary research questions:

1) Does a long or short RR window prior to therapy lead to better retrospective discrimination?
    a. Short 16 RR windows were far superior to longer windows of variable RR length ranging from 21-40 RRs when completing retrospective discrimination.
2) Does arrythmia type and/or manufacturer influence discrimination?
    a. There was no distinction between discrimination of VT or VF from AF. Inappropriate BSci cases were further from the appropriate therapy group than inappropriate Medtronic cases.
3) Does separating the data by manufacturer into two separate datasets improve discrimination?
    a. Separating the data by manufacturer improved discrimination, but ideally all cases should be discriminable independent of manufacturer.
4) Do the therapy scores discriminate appropriate and inappropriate therapy?
    a. Linear combination scores performed similarly to the PCA strategy used in answering research questions 1-3.
5) Do the retrospective strategies work for real time decision making?
    a. Both PCA and therapy scores show promise for use in real time therapy decision making, but the therapy scores outperformed PCA.

From these answers, three major conclusions were made:

1. Both PCA and the therapy scores show promise in discriminating between appropriate and inappropriate therapies retrospectively across Medtronic and BSci single chamber TV and S-ICD devices.
2. When completing retrospective analysis, short windows just prior to therapy provide clear visual separation between the inappropriate and appropriate therapy groups, while long windows do not.
3. Real time therapy decision making is feasible with both PCA and the therapy scores, but the therapy scores in conjunction with the ATZ show more promise.

4.2 – Limitations

The major limiting factor of this entire experiment was the small and unbalanced dataset. Recent clinical studies have nearly 10x the number of appropriate therapies and 10x to 100x the number of inappropriate therapies in their datasets [4, 7]. To draw conclusions of similar significance, the dataset would need to be at least the same size, if not a larger size. Moreover, more inappropriate events from S-ICD devices are needed, especially since most of the inappropriate therapy in S-ICD devices is caused by oversensing, not SVTs.

A limitation of the "pseudo" real time analysis was that many of the events were already in arrythmia at the start of the analysis in Window 1. Ideally, the signals would all have started in NSR or other baseline rhythm and then transitioned into VT/VF. Otherwise, it is unknown if some of the patient's baseline rhythms would fall within the ATZ and cause the delivery of an inappropriate therapy. Lastly, this experiment only looked at Medtronic and BSci devices. Including other manufacturers such as Biotronik would make the experiment more robust.

4.3 – Future Directions

An algorithm utilizing the therapy score and ATZ should be investigated first in a larger retrospective study with a dataset tenfold larger. Machine learning could also be used in this study to optimize the ATZ. The next step would be to program this into actual devices and then complete a clinical study in conjunction with the manufacturers. The algorithm could also be used retrospectively to adjudicate events and relieve

Cardiologists of this duty. However, it must first be validated on a larger dataset to ensure it has high accuracy prior to using it for this purpose.

Bibliography

[1] D. Friedman, C. Parzynski and P. Varosy, "Trends and In-Hospital Outcomes Associated With Adoption of the Subcutaneous Implantable Cardioverter Defibrillator in the United States," *JAMA Cardiol.,* pp. 900-911, 2016.

[2] N. Ghosh, I. Mangat, S. O'Donnell, A. Pinter, V. Korley, C. Lane and P. Dorian, "Outcomes in heart failure patients referred for consideration fo implantable cardioverter defibrillator for primary prophylaxis of sudden cardiac death: What are the risks of waiting?," *Can J Cardiol.,* pp. e342-e346, 2009.

[3] B. Schrage, A. Uijl, L. Benson, D. Westermann, M. Stahlberg, D. Stolfo, C. Linde, F. Braunschweig and G. Savarese, "Association Between Use of Primary-Prevention Implantable Cardioverter-Defibrillators and Mortality in Patients With Heart Failure," *Circulation,* pp. 1530-1539, 2019.

[4] A. Moss, C. Schuger, C. Beck, M. Brown, D. Cannom, J. Daubert, ... and W. Zareba, "Reduction in Inappropriate Therapy and Mortality through ICD Programming," *N ENGL J MED,* pp. 2275-2283, 2012.

[5] C. Morillo, A. Banerjee, P. Perel, D. Wood and X. Jouven, "Atrial fibrillation: the current epidemic," *J Geriatr Cardiol.,* pp. 195-203, 2017.

[6] M. Gold, P. Lambiase, M. El-Chami, K. Reinoud, J. Aasbo, M. Bongiorni, ... and L. Boersma, "Primary Results From the Understanding Outcomes With the S-ICD in Primary Prevention Patients With Low Ejection Fraction (UNTOUCHED) Trial," *Circulation,* pp. 7-17, 2021.

[7] A. Auricchio, E. Schloss, T. Kurita, A. Meijer, B. Gerritse, S. Zweibel, ... and L. Sterns, "Low inappropriate shock rates in patients with single- and dual/triple-chamber implantable cardioverter-defibrillators using a novel suite of detection algorithms: PainFree SST trial primary results," *Heart Rhythm,* 2015.

[8] Boston Scientific Corporation, "Emblem MRI S-ICD System - Clinical Data," 2020. [Online].

[9] R. Passman, H. Subacius, B. Ruo, A. Schaechter, A. Howard, S. Sears and A. Kadish, "Implantale Cardioverter Defibrillators and Quality of Life: Results From the Defibrillators in Nonischemic Cardiomyopathy Treatment Evaluation Study," *Arch Intern Med.,* pp. 2226-2232, 2007.

[10] E. Schron, D. Exner, Q. Yao, L. Jenkins, J. Steinberg, J. Cook, ... and J. Powell, "Quality of life in the antiarrhythmics versus implantable defibrillators trial: impact of therapy and influence of adverse symptoms and defibrillator shocks," *Circulation,* pp. 589-594, 2002.

[11] F. Mansour and P. Khairy, "Ventricular tachycardia detection with the subcutaneous implantable cardioverter-defibrillator: Tracing the cause of an absent tracing," *HeartRhythm Case Rep.,* pp. 77-80, 2020.

[12] J. Saiz-Vivo, V. Corino, M. Melis and L. Mainardi, "Unsupervised Classification of Atrial Fibrillation Triggers Using Heart Rate Variability Features Extracted from Implantable Cardiac Monitor Data," in *Annu Int Conf IEEE Eng Med Biol Soc.*, 2020.

[13] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson and Y. Zhang, "Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy," *Biomed Eng Online,* 2014.

[14] S. Arunachalam, S. Kapa, S. Mulpuru, P. Friedman and E. G. Tolkacheva, "Improved Multiscale Entropy Technique with Nearest-Neighbor Moving-Average Kernel for Nonlinear and Nonstationart Short-Time Biomedical Signal Analysis," *Journal of Healthcare Engineering,* 2018.

[15] V. Ravikumar, E. Annoni, S. Mulpuru, H. Roukoz and E. Tolkacheva, "Evaluation of multiscale frequency approach for visualizing rotors in patients with atrial fibrillation," *Proceedings of IEEE EMBC,* 2018.

[16] S. Arunachalam, E. Annoni, S. Mulpuru, P. Friedman and E. Tolkacheva, "Novel Multiscale Frequency Approach to Identify the Pivot Point of the Rotor," *Journal of Medical Devices,* vol. 10, 2016.

[17] Mathworks, "Kurtosis," 2020. [Online]. Available: https://www.mathworks.com/help/stats/kurtosis.html. [Accessed 06 05 2020].

[18] G. Lewis and M. Gold, "Clinical experience with subcutaenous implantable cardioverter-defibrillators," *Nature Reviews Cardiology,* pp. 398-405, 2015.

Appendix

A.1 – Digitization

1. Screenshot the desired EGM signal.
2. In the MacOS preview, adjust the threshold of the cropped image to remove dotted lines and other noise from the image so only the EGM signal is left. MATLAB can also be used for this.
3. Save as a 300 dpi image with size adjustment making sure to not resample the image.
4. Upload the image to MATLAB.
5. Using a custom written MATLAB script, set all pixels in the image with a grayscale value less than 200 to 0. The only remaining pixels with a value other than 0 should be the pixelated trace of the EGM signal.
6. Plot the pixelated trace and fit a smoothing spline to the pixelated trace.
7. Using a discrete index set by your desired sampling frequency and length of the EGM trace in time (taken from the case reports), sample the smoothing spline fit to get the digitized signal.
8. Plot the final digitized signal and the pixelated trace to ensure high fidelity digitization.
9. Save the digitized signal in ASCII format for analysis with MATLAB.

A.2 – Detailed description of Nonlinear-based Metrics

MSE & SE

Entropy measures such as multi-scale entropy (MSE) are typically utilized to quantify the complexity of a signal through quantification of a signals "uncertainty." Commonly used entropy measures for biomedical signals includes Shannon Entropy (SE), Sample Entropy, etc. In fact, SE has shown decent prognosis in identifying AF on ECG data [13]. Thus, SE may work well to identify AF when compared with VF/VT. However, standard entropy based measures such as SE may fail when the data becomes non-stationary such as with EGM recordings. MSE was designed to overcome this limitation by incorporating information from the "past" and "future" in the form of a moving average kernel. Because the heart rhythm becomes more complex during VF/VT

compared with non-VF/VT rhythms, it was hypothesized that the improved MSE will significantly increase once the heart enters VF/VT.

The improved MSE utilizes a nearest-neighbor moving-average kernel. This measure has previously been used to accurately discriminate atrial fibrillation [14]. A full description of the technique can be found in Ref. [14]. For this MSE, a scale factor τ must be chosen for the nearest-neighbor moving-average kernel as τ controls the number of samples averaged over. A scale factor of $\tau = 10$ was chosen for this research as it has been shown that the optimal τ for rotor pivot point identification is 10 in unpublished work by the Talkachova lab. Values of the delay factor, δ, and length of matching segments, m, was left as 1 and 2 as was previously suggested for atrial fibrillation discrimination [14].

MSF

In this study, an improved multiscale frequency measure is utilized. It has previously been reported to accurately identify the pivot point of rotors [15] [16]. A full description of the technique can be found in Ref. [16]. The improved MSF combines local estimates of instantaneous frequency over many time scales. This helps account for the nonstationary nature of the heart which makes common frequency based approaches less useful for analysis of EGMs. The center frequency for the 8 log-Gabor filters used in this method were chosen to span a physiological range for the human heart. Thus, it is hypothesized that as the patient enters VF/VT their EGM will change in frequency characteristics and the MSF value will be different then in the case of non VF/VT such as AF, NSR or oversensing.

Kurtosis

Sample Kurtosis (simplified to kurtosis in this study) is the third feature measured in this study. Kurtosis is a higher order statistical measure known as the fourth standardized moment of the sample distribution. It measures the combined weight of the tails of the sample distribution in comparison to the weight of the center of the sample distribution. The kurtosis of a normal distribution is 3. Distributions with larger tail to center ratios than a normal distribution have kurtosis greater than 3 and those with smaller ratios have kurtosis less than 3. The definition of kurtosis utilized in this study is given by MATLAB in Ref. [17] where kurtosis is assumed as biased with a flag of 1.

Kurtosis should note changes in the morphology of signals. Thus, it was expected that the distinct morphological changes in a VF/VT signal compared to a non-VF/VT signal would cause kurtosis to be significantly different between VF/VT and non-VF/VT rhythm.

A.3 – Description of RR-based metrics

1. MeanRR
    a. The MeanRR is simply the mean value of all of the RR values.
2. pNN50
    a. pNN50 stands for the percentage of NNs (differences between successive RR intervals) greater than 50 ms. The larger pNN50 is, the more variability in the RR intervals.
3. RMS
    a. The root mean squared value of the NNs.
4. STD
    a. The standard deviation of the NNs.
5. SD1
    a. The variance between the sum of the current RR intervals and the next RR intervals divided by the square root of 2. Typically associated with short-term RR interval variability.
6. SD2
    a. The variance between the difference of the current RR intervals and the next RR intervals divided by the square root of 2. Typically associated with long-term RR interval variability.
7. SD1/SD2
    a. The ratio between SD1 and SD2.